

UNCLASSIFIED

AD 410352

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA



UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

H-63-4-3

# Research Report

PEER RATING VALIDITY AS A FUNCTION OF RATER INTELLIGENCE AND  
RATING SCORE RECEIVED

Richard E. Doll

Bureau of Medicine and Surgery  
Project MR005.13-5001  
Subtask 1 Report No. 24

CATALOGED BY DDC

AS AD No. 410352

Approved by

Captain Ashton Graybiel, MC, USN  
Director of Research

Released by

Captain Clifford P. Phoebus, MC, USN  
Commanding Officer

15 March 1963

U. S. NAVAL SCHOOL OF AVIATION MEDICINE  
U. S. NAVAL AVIATION MEDICAL CENTER  
PENSACOLA, FLORIDA

410352

## SUMMARY PAGE

### THE PROBLEM

Peer Ratings used in the Naval Air Training Program have proved quite useful in predicting subsequent failures. This study attempts to determine the relationship of two rater "characteristics," intelligence and Peer Rating score received, to the validity of the ratings given.

### FINDINGS

Results from three analytic approaches to the records of 548 cadets demonstrate that when dealing with a population having generally above average intelligence there is little reason to take into consideration rater intelligence when concerned with the validity of the ratings he gives. This is also true for the Peer Rating score received by the rater.

<p>Doll, R.E.</p> <p>1963</p> <p>PEER RATING VALIDITY AS A FUNCTION OF RATER INTELLIGENCE AND RATING SCORE RECEIVED. Project MR005.13-5001 Subtask 1, Report No. 24. Pensacola, Fla.: Naval School of Aviation Medicine, 15 March.</p> <p>Peer Ratings used in the Naval Air Training Program have proved quite useful in predicting subsequent failures. This study attempts to determine the relationship of two rater "characteristics," intelligence and Peer Rating score received, to the validity of his ratings.</p> <p>Results from three analytic approaches to the records of 548 cadets demonstrate that when dealing with a population of above average intelligence there is little reason to take into consideration rater intelligence when concerned with the validity of the ratings he gives. This is also true for the Peer Rating score received by the rater.</p> <p>Sociometrics Aviation training- Performance</p> <p>Psychology</p> <p>L. C. Subj. Head.</p>	<p>Doll, R.E.</p> <p>1963</p> <p>PEER RATING VALIDITY AS A FUNCTION OF RATER INTELLIGENCE AND RATING SCORE RECEIVED. Project MR005.13-5001 Subtask 1, Report No. 24. Pensacola, Fla.: Naval School of Aviation Medicine, 15 March.</p> <p>Peer Ratings used in the Naval Air Training Program have proved quite useful in predicting subsequent failures. This study attempts to determine the relationship of two rater "characteristics," intelligence and Peer Rating score received, to the validity of his ratings.</p> <p>Results from three analytic approaches to the records of 548 cadets demonstrate that when dealing with a population of above average intelligence there is little reason to take into consideration rater intelligence when concerned with the validity of the ratings he gives. This is also true for the Peer Rating score received by the rater.</p> <p>Sociometrics Aviation training- Performance</p> <p>Psychology</p> <p>L. C. Subj. Head.</p>
<p>Doll, R.E.</p> <p>1963</p> <p>PEER RATING VALIDITY AS A FUNCTION OF RATER INTELLIGENCE AND RATING SCORE RECEIVED. Project MR005.13-5001 Subtask 1, Report No. 24. Pensacola, Fla.: Naval School of Aviation Medicine, 15 March.</p> <p>Peer Ratings used in the Naval Air Training Program have proved quite useful in predicting subsequent failures. This study attempts to determine the relationship of two rater "characteristics," intelligence and Peer Rating score received, to the validity of his ratings.</p> <p>Results from three analytic approaches to the records of 548 cadets demonstrate that when dealing with a population of above average intelligence there is little reason to take into consideration rater intelligence when concerned with the validity of the ratings he gives. This is also true for the Peer Rating score received by the rater.</p> <p>Sociometrics Aviation training- Performance</p> <p>Psychology</p> <p>L. C. Subj. Head.</p>	<p>Doll, R.E.</p> <p>1963</p> <p>PEER RATING VALIDITY AS A FUNCTION OF RATER INTELLIGENCE AND RATING SCORE RECEIVED. Project MR005.13-5001 Subtask 1, Report No. 24. Pensacola, Fla.: Naval School of Aviation Medicine, 15 March.</p> <p>Peer Ratings used in the Naval Air Training Program have proved quite useful in predicting subsequent failures. This study attempts to determine the relationship of two rater "characteristics," intelligence and Peer Rating score received, to the validity of his ratings.</p> <p>Results from three analytic approaches to the records of 548 cadets demonstrate that when dealing with a population of above average intelligence there is little reason to take into consideration rater intelligence when concerned with the validity of the ratings he gives. This is also true for the Peer Rating score received by the rater.</p> <p>Sociometrics Aviation training- Performance</p> <p>Psychology</p> <p>L. C. Subj. Head.</p>

## INTRODUCTION

Peer Ratings, i.e., evaluations of the individual in a group by one or more other individuals in that group, have proved to be useful instruments. Such ratings, even though made by untrained and relatively unsophisticated observers, have been shown to be good predictors of relative success or failure in several areas of endeavor. Studies have indicated that such ratings have substantial validity in predicting flight failure (2), officer efficiency rating (10), military grades in Officer Candidate School (9), leadership performance in combat (8), and on-the-job performance (6).

Such Peer Ratings are among the measures used in the Naval Air Training Program to appraise the potential of individual cadets (i.e., Aviation Officer Candidates, Naval Aviation Cadets, and Marine Aviation Cadets). During the eighth week of training each man in a class is asked to name the three most promising prospective officers and the three least promising in his class. It has been shown (4,5) that these ratings typically have a biserial correlation of about .35 with subsequent completion or failure to complete the training program and that, when combined with other measures, they have considerable administrative usefulness.

It can be expected that student raters differ in the validity of their ratings. It can also be demonstrated that they vary on many other measures. This study attempts to determine whether the differences in the validities of ratings given are related to differences among cadets on two other variables, intelligence and on the rater's own Peer Rating score.

Browning, et al. (3), in one of a series of U.S. Army studies investigating rating methodology, attempted to answer the same question. Their results showed a moderate positive relationship between raters' intelligence scores and the validity of their ratings. There was also a very slight positive relationship between the Peer Rating scores received by the raters and validity of their ratings. However, a serious weakness in this series of studies has been pointed out (7), in that the criterion used (a proficiency rating completed earlier by the same people serving as subjects) suffered from both rater contamination and technique contamination. The Naval Air Training Program, however, affords an independent and ultimate criterion (i.e., complete/fail to complete), making it possible to repeat Browning's study without the contaminating factors.

## PROCEDURE

The data consisted of the records for 548 cadets from 30 pre-flight school classes who entered the program during 1959. All the cadets had at least two years of college education.

## MEASURES

The scoring of Peer Ratings varies with the format of the ratings. The format used in the Naval Air Training Command calls for a ratee to receive a +3 every time he is named most promising, +2 for second most promising, +1 for third most promising, a -3 for least promising, et cetera. These values are then algebraically summed and divided by the number of raters in the class. If the cadet ratee has not been nominated either high or low, he received a 0. These quotients are converted to a standard score through conversion tables based upon norms for past classes.

Intelligence is operationally defined in this study as being the score on the Aviation Qualification Test (AQT). This test correlates .70 with the American Council of Education Psychological Examination (ACE) and .71 with the Wonderlic Personnel Test. The mean scores of cadets on both the ACE and Wonderlic Personnel Test are significantly higher than the mean scores obtained by the general population (1). This, plus the two years of college prerequisite for selection, allows one to assume that the population used in this study is of above average intelligence. The second independent variable, Peer Rating score, has already been discussed.

## METHODS

Three analytic approaches were utilized in this study. The first approach consisted of giving each rater a score based on the accuracy of the ratings he gave his peers. For each of the three cadets whom the rater rated as being most promising, a +1 was given if the ratee completed the program; however, -1 was given if the ratee failed to complete. Also a +1 was given for each peer who was rated as being least promising and who failed to complete the program. Conversely, for each low rated cadet who completed the program the cadet rater received a -1. The cadet rater's score was the algebraic sum of the pluses and minuses (with a constant of +10 added to avoid negative scores). This score made up the peer rating accuracy (PRA) score reflecting, as it did, the validity of the ratings given by the cadet. In the first approach the rater's PRA score was correlated with his AQT score and his Peer Rating score in order to determine the relationships between PRA and intelligence and between PRA and rating score received.

In the second approach, the procedure used by Browning, *et al.*, was adopted. Instead of treating each rater individually, the total group was divided into thirds, (high, medium, and low) on each of the two independent variables, AQT score and Peer Rating score. Thus each class was roughly divided into thirds. With each third treated as though it were a class the Peer Ratings were scored following the usual format. Each cadet in the class received three Peer Rating scores, one assigned by each third. A biserial correlation was then computed between the recorded Peer Rating scores assigned by each of the upper, middle, and lower groups and the criterion of complete/fail to complete the training program.

The third approach divided the raters into greater than  $\pm 1$  S.D. on AQT score and Peer Rating received score. The differences between the PRA means of these extreme groups were tested.

## RESULTS AND DISCUSSION

Examination of Table I makes it quite evident that there is no relationship between rater intelligence (AQT score), Peer Rating score, and the validity of the rater's ratings when treated on an individual basis.

Table I

Correlations\* Between Rater AQT Score, Peer Rating Score, and the Criterion (PRA) When Treated Individually

Variables	Criterion (PRA)
AQT	-.01
Peer Rating Score	.04

\* Pearson Product-Moment Correlation

The second approach, however, yields somewhat different results, as shown in Table II. There exists a slight positive relationship between the intelligence level of the three groups as measured by the AQT and the validity of the pooled ratings of the members of the respective groups. The three groupings according to Peer Rating score, on the other hand, have a relationship that tends to be U-shaped, with the upper third demonstrating a superior rating performance. Aside from this relationship the results are in the expected direction and conform reasonably well with the results published by Browning, *et al.*

Table II

Validity Coefficients\* of Peer Ratings# by Groups of Raters  
Falling into Upper, Middle, And Lower Thirds on AQT Score and  
Peer Rating Score When Criterion is Complete/Fail to Complete

Variable	Upper Third	Middle Third	Lower Third
AQT	.29	.25	.25
Peer Rating Score	.33	.23	.27

\*Biserial correlation

#The correlation between Peer Ratings for the total group and the criterion of complete/fail to complete was .37. The drop in correlations when divided into thirds probably was due to curtailment of range.



Table III shows the results of the third approach which compared the mean PRA scores of just the extreme (i.e., greater than  $\pm 1$  S.D.) groups on AQT and Peer Rating score received. It is readily apparent that no significant differences exist.

Table III

Comparison of PRA Mean Scores and Standard Deviation for Greater Than  $\pm 1$  S.D. on AQT Score and Peer Rating Score

	Greater than +1 S.D.			Greater than -1 S.D.			CR*	p
	N	Mean PRA	S.D.	N	Mean PRA	S.D.		
AQT	116	11.63	2.77	134	11.77	2.25	.70	N.S.
Peer Rating Score	75	12.01	2.20	67	11.58	2.45	.53	N.S.

\*Critical Ratio test of significance for difference between means

### CONCLUSION

The results shown in Table II conform reasonably well with the results published by Browning, et al. In both studies a slight positive relationship was found between the validity of a rater's peer ratings and the two variables, rater intelligence and Peer Rating score received by the rater. The other two approaches (Tables I and III), however, show no apparent relationship between these variables. Since the second procedure in this study was the same as that used by Browning, et al., and both yielded similar results, one is tempted to conclude similarity of results is in part due to similarity of analysis. In the light of this reproduction of results it could also be concluded that procedure two is the more sensitive or valid analysis due perhaps to the "pooling of judgments." Any conclusions derived from Table II, however, are weak because of the statistical difficulty in testing for the differences between biserial correlations.

For the most part the differences are slight. Even should the differences be statistically significant, it is apparent that they are not of the magnitude to be considered practically significant. If the relationships were of such a magnitude, surely it would have come out in the other two analytic approaches.

Results therefore seem to indicate that, when dealing with individuals within the intelligence range used in this study, there is little practical reason to take into consideration rater intelligence when concerned with the validity of the ratings he gave, at least for this criterion. This is also true for the Peer Rating score received by the rater.

## REFERENCES

1. Ambler, R.K., Differences between aviation officer candidates and naval aviation cadets on three tests of mental ability. Project MR005.13-3003 Subtask 1, Report No. 13. Pensacola, Fla.: Naval School of Aviation Medicine, 1956.
2. Berkshire, J.R., and Nelson, P.D., Leadership Peer Ratings related to subsequent proficiency in training and in the fleet. Special Report 58-20. Pensacola, Fla.: Naval School of Aviation Medicine, 1958.
3. Browning, R.C., Campbell, J.T., Birnbaum, A.H., Haggerty, H.R., and Scheider, D.E., A study of officer rating methodology. X. Effects of selected rater characteristics on validity of ratings. PRS Report 909. Washington, D.C.: The Adjutant General's Office, 1952.
4. Doll, R.E., Officer Peer Ratings as a predictor of failure to complete flight training. Special Report 62-2. Pensacola, Fla.: Naval School of Aviation Medicine, 1962.
5. Doll, R.E., and Berkshire, J.R., The validity of the officer-like-quality measures used in the U.S. Naval School, Pre-Flight. Special Report No. 61-6. Pensacola, Fla.: Naval School of Aviation Medicine, 1961.
6. Fuchs, E.R., Woods, I.A., and Harper, B.P., Prediction of job success in eight career ladders. PRB Report 997. Washington, D.C.: The Adjutant General's Office, 1953.
7. Karcher, E.K., Jr., Campbell, J.T., Falk, G.H., and Haggarty, H.R., A study of officer rating methodology. VI. Independence of criterion measures from predictor variables. PRS Report 905. Washington, D.C.: The Adjutant General's Office, 1952.
8. Lindzey, G., and Borgatta, E.F., Sociometric measurement. In Lindzey, G. (Ed.), Handbook of Social Psychology. Vol. 1: Theory and method. Cambridge, Mass.: Addison-Wesley, 1954.
9. Suci, G.J., and Vallance, T.R., An analysis of Peer Ratings: II. Their validity as predictors of military aptitude and other measures in Naval Officer Candidate School. Technical Bulletin 54-10. Washington, D. C.: Bureau of Naval Personnel, 1954.
10. Wherry, R.J., and Fryer, D.H., Buddy ratings: Popularity contest or leadership criteria? Personn. Psychol., II: 147-159, 1949.